



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB III

### METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM

#### 3.1 Metodologi Penelitian

Dalam penelitian implementasi algoritma *Gaussian Naive Bayes* pada deteksi sentimen *cyberbullying* di media sosial instagram, metode penelitian akan dibagi menjadi 6 tahap, yaitu studi literatur, pengumpulan data, analisa masalah, perancangan sistem, mengimplementasi algoritma yang digunakan, dan prosedur penelitian.

##### 3.1.1 Studi Literatur

Pada tahap ini dilakukan pembelajaran terhadap teori-teori dengan mencari jurnal, *paper*, serta artikel terpercaya melalui Internet yang berhubungan dengan *cyberbullying*, *Gaussian Naive Bayes*, *N-Gram*, dan *TF-IDF*.

##### 3.1.2 Pengumpulan Data

Melakukan *scrapping comment* untuk mendapatkan data komentar pengguna instagram pada suatu unggahan dari komentar unggahan instagram dari *instagrammer* Indonesia yang sudah *verified*. Selanjutnya dilakukan labelisasi data yang dibantu oleh pakar bahasa Indonesia.

### **3.1.3 Analisa Masalah**

Pada tahap ini dilakukan analisa informasi yang telah diperoleh untuk mendapat pemahaman terhadap masalah dan metode yang digunakan untuk penyelesaian masalah yang telah diangkat.

### **3.1.4 Perancangan Sistem**

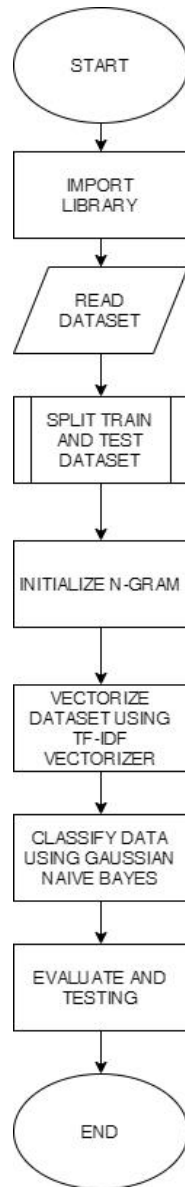
Pada tahap ini dilakukan perancangan sistem yang akan dibuat sesuai dengan pembahasan sebelumnya. Perancangan yang dimaksud merupakan pembuatan *flowchart* untuk mempermudah alur program yang telah dikembangkan.

### **3.1.5 Implementasi Algoritma yang Digunakan**

Pada tahap ini akan dilakukan tahap mengimplementasi algoritma yang digunakan pada sistem klasifikasi komentar instagram untuk *sentiment analysis cyberbullying*. Selain itu dilakukan juga implementasi algoritma pendukung seperti *N-gram*, dan TF-IDF.

### 3.1.6 Prosedur Penelitian

Penelitian akan dilakukan dalam beberapa tahap sebagai berikut.



Gambar 3.1.Flowchart penelitian

#### A. *Import Libraries*

Melakukan *import library* yang dibutuhkan dalam penelitian yang dilakukan.

## B. Read Dataset

Pada tahap ini, *dataset* yang diolah akan dimasukkan yang kemudian dianalisa.

## C. Split Train and Test Dataset

Tahapan ini akan membagi data ke dalam data latih dan data uji dengan rasio yang telah ditentukan sesuai skenario.

### C.1 Train set

*Train set* merupakan data yang digunakan untuk melatih model *machine learning* sebagai contoh agar dapat bekerja sesuai parameter yang diberikan.

### C.2 Test set

*Test set* merupakan data yang digunakan untuk memberikan evaluasi model setelah dilakukan *training*.

## D. Initialize N-gram

Pada bagian ini akan dilakukan proses *n-gram* menggunakan angka minimal 3 dan maksimal 3 dengan menggunakan *analyzer char\_wb*.

## E. Vectorize Data Using Tf-Idf Vectorizer

Model data yang belum dilakukan *training* akan ditanamkan ke dalam fungsi *TfidfVectorizer* yang kemudian digunakan sebagai *vectorizer* terhadap setiap input kalimat.

## F. Classify Data Using Gaussian Naive Bayes

Tahap ini akan melakukan *training data* yang telah dipisah sebelumnya. *Test set* dimasukkan ke dalam model hasil latih dengan parameter yang ditentukan.

Klasifikasi akan dilakukan saat model yang dibuat telah dievaluasi terhadap *test set* dan *train set* yang ada. Klasifikasi dilakukan dengan algoritma *Gaussian Naive Bayes*.

### G. Evaluate and Testing

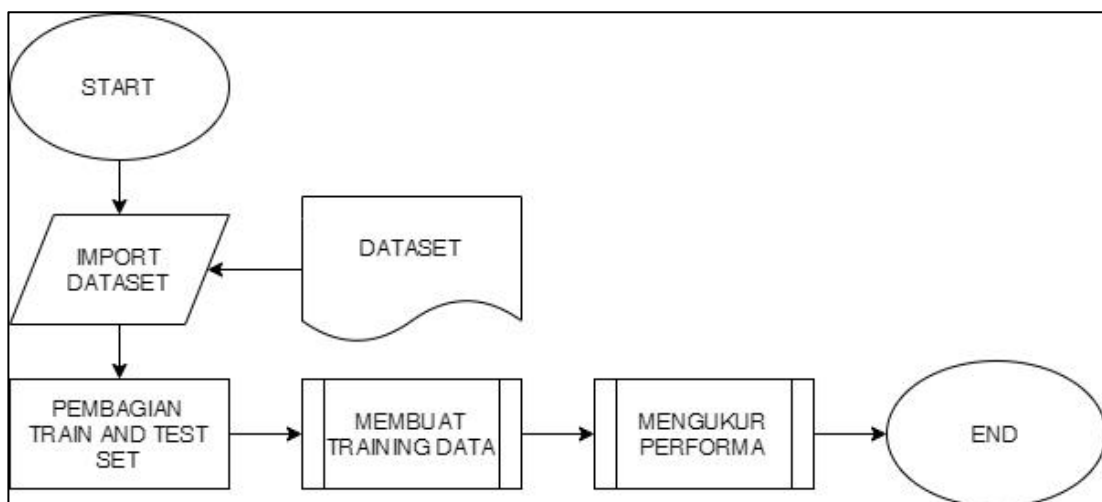
Tahap terakhir dengan melakukan evaluasi performa untuk mendapatkan nilai akurasi, *precision*, dan *recall*. Kemudian nilai yang didapat akan digunakan untuk melakukan perhitungan F1. Proses Testing dilakukan dengan memasukkan suatu kalimat dan melihat *outputnya*. Jika *output* “1” maka kalimat tersebut termasuk *bully* dan jika ‘0’ kalimat tersebut termasuk tidak *bully*.

## 3.2 Perancangan Aplikasi

Perancangan aplikasi dilakukan dengan menggunakan *flowchart*.

### 3.2.1 Flowchart

*Flowchart* merupakan diagram yang menunjukkan alur kerja program. Dalam penelitian yang dilakukan, alur utama dalam penelitian digambarkan pada Gambar 3.2 sebagai berikut.



Gambar 3.2 Flowchart pembuatan model

Gambar 3.2 menunjukkan aktivitas *training data* untuk mendapatkan model yang diharapkan. Langkah awal yang dilakukan dalam penelitian adalah dengan melakukan *import dataset* yang telah dianotasikan pakar bahasa Indonesia. *Dataset* yang digunakan berjumlah 400 data. Langkah pertama yang dilakukan dengan membagi data tersebut menjadi data latih dan data uji. Berikutnya dilakukan proses pemodelan data latih dan pengujian model yang dibuat.

Berikutnya dijelaskan alur pemodelan data latih. Alur pemodelan data latih ditunjukkan pada Gambar 3.3 sebagai berikut.

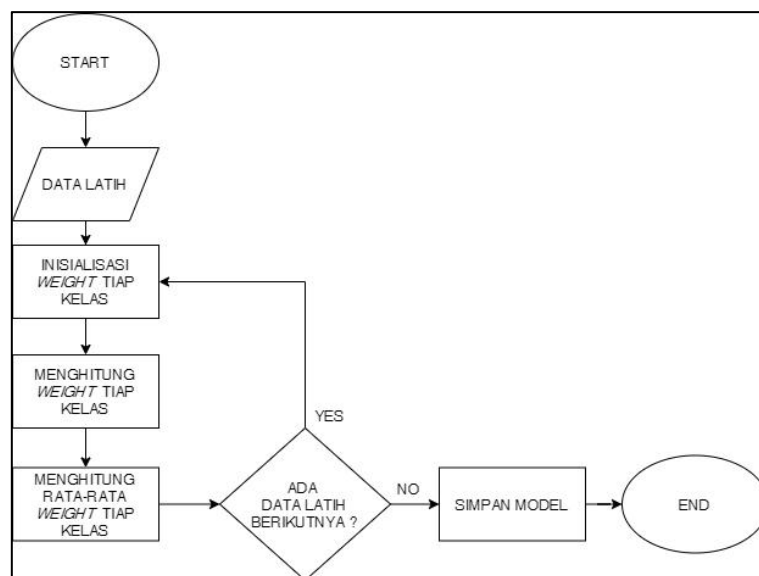


Gambar 3.3 Flowchart Pemodelan data latih

Pada Gambar 3.3, dijelaskan pembuatan *training data* untuk kemudian diproses. Langkah awal yang dilakukan dengan inisialisasi *N-gram*. Pada penelitian kali ini digunakan *Tri-gram* yang berarti setiap 3 karakter dalam suatu kalimat akan

dipisah. Selanjutnya data tersebut diubah menjadi vektor *TF-IDF*. Langkah terakhir mulai membuat model *Gaussian Naive Bayes* dari data yang sudah menjadi vektor oleh *TD-IDF*.

Bagian berikutnya dijelaskan alur pembuatan model *gaussian naive bayes*. *Flowchart* pembuatan model *gaussian naive bayes* ditampilkan pada Gambar 3.4 berikut ini.

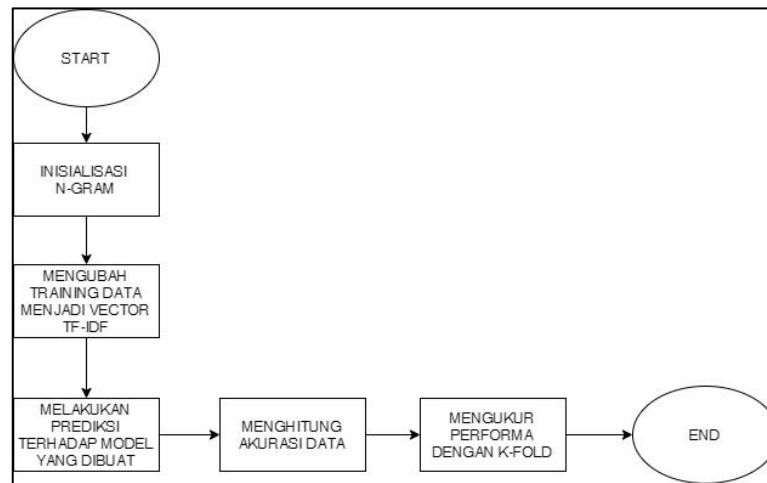


Gambar 3.4 Flowchart Gaussian Naive Bayes

Pada Gambar 3.4, dijelaskan proses pembuatan model *gaussian Naive Bayes* dengan menggunakan data latih yang telah dibuat. Langkah awal dengan inisialisasi bobot tiap kelas. Berikutnya dilakukan pengulangan terhadap data latih tersebut dan melakukan penghitungan bobot setiap pengulangan. Begitu pengulangan berakhir model yang telah dibuat akan disimpan dan akan dilakukan proses pengujian.

Bagian terakhir menjelaskan proses pengujian performa model. Perancangan alur pengujian performa model ditampilkan pada Gambar 3.5 menggunakan *flowchart* sebagai berikut.





Gambar 3.5 Flowchart pengujian model

Pada Gambar 3.5, dijelaskan proses pengujian performa model yang telah dibuat. Langkah pertama yang dilakukan dengan inisialisasi *N-Gram*. Setelah itu dilanjutkan dengan mengubah data tersebut menjadi vektor *TF-IDF*. Begitu selesai, mulai dilakukan proses prediksi dengan menggunakan model *Gaussian Naive Bayes* yang telah dibuat sebelumnya. Dari hasil prediksi tersebut, dilakukan proses penghitungan akurasi data. Setelah itu dilakukan *K-Fold* untuk mengukur performa *Gaussian Naive Bayes*.